

## TELEPRESENCE USING PANORAMIC IMAGING AND DIRECTIONAL SOUND

### CROSS-REFERENCE TO RELATED APPLICATION

This application claims the benefit of U.S. Provisional Patent Application Serial  
5 No. 60/180,620 filed February 7, 2000.

### FIELD OF THE INVENTION

The present invention relates to panoramic imaging, and more particularly  
relates to the use of panoramic visual images in combination with directional sound to provide  
an emersive imaging experience.

### BACKGROUND INFORMATION

10 Panoramic imagery is able to capture a large azimuth view with a significant  
elevation angle. In some cases, the view is achieved through the use of wide angle optics such  
as fish-eye lenses. In other cases, it is achieved through the use of a combination of mirrors  
and lenses. Alternatively, the view may be developed by rotating an imaging sensor so as to  
15 achieve a panorama. The panoramic view can be composed of still images or, in cases where  
the images are taken at high frequencies, the sequence can be interpreted as animation. Wide  
angles associated with panoramic imagery can cause the image to appear warped, i.e., the  
image does not correspond to a natural human view. This imagery can be unwarped by  
various means including software to display a natural view.

20 U.S. Patent No. 5,771,041 to Small discloses a system for producing directional  
sound in computer-based virtual environments. This reference describes how sounds can be  
played back as opposed to how they are collected. Moreover, these sounds are recorded only  
as point sources and no provision is available for directional capture of sound or capture of  
diffuse sounds.

25 While systems have been proposed in which panoramic images can be created in  
computer generated environments, such as with three dimensional models, the present  
invention relates to photographed still or video imagery that is combined with directional sound  
to provide telepresence.

## SUMMARY OF THE INVENTION

Visual images and sound are very important to provide a complete sense of place. The present telepresence system conveys not only visual information but also audio information to improve the realism of the experience.

5 An aspect of the present invention is to provide an imaging system comprising a panoramic visual images display, and an associated directional sound playback device.

Another aspect of the present invention is to provide an imaging system and method for providing a panoramic visual images, and for providing directional sound associated with the panoramic visual images.

10 A further aspect of the present invention is to provide an image recording system comprising a panoramic visual images recording device, and an associated directional sound capturing device.

Another aspect of the present invention is to provide an image recording system and method for capturing panoramic visual images, and for capturing directional sound associated with the panoramic visual images.

15 A further aspect of the present invention is to provide an image recording and playback system comprising a panoramic visual images recording device, an associated directional sound capturing device, a panoramic visual images display, and an associated directional sound playback device.

20 Another aspect of the present invention is to provide an image recording and playback system and method for capturing panoramic visual images, for capturing directional sound associated with the panoramic visual images, for providing panoramic visual images, and for providing directional sound associated with the panoramic visual images.

25 These and other aspects of the present invention will be more apparent from the following description.

## BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a schematic diagram illustrating a system for producing panoramic images.

Fig. 2 is a raw image from a panospheric camera.

Fig. 3 is the image from Fig. 2 displayed as a rectangular image using a projection onto a cylindrical surface.

Fig. 4 is a schematic illustration of a single panoramic camera with multiple microphones which are used to record sound from one or more sound sources.

Fig. 5 schematically illustrates how the loudness of a sound from a particular sound source depends upon its location with respect to the current viewing direction.

Fig. 6 is the rectangular projected image of Fig. 3, illustrating an angle between the viewing direction (the center of the selected view) and the reference frame of the camera.

Fig. 7 schematically illustrates sound that is recreated when microphones are not at the optical center of a panoramic device.

Fig. 8 schematically illustrates multiple panoramic cameras in combination with multiple microphones which are used to record sound from one or more sound sources.

### DETAILED DESCRIPTION

The present invention combines panoramic visual images and directional sound. The panoramic visual images can comprise one or more individual still images, or a sequence of images such as a video stream. An aspect of the invention is that sound recorded with more than one recording device can be played back in conjunction with panoramic images based on a particular viewing direction. Upon playback, directional sound associated with the particular view convey a life-like experience.

As used herein, the term "panoramic visual images" means wide angle images taken from a field of view of from about 60° to 360°, typically from about 90° to 360°. Preferably, the panoramic visual images comprise a field of view of from about 180° to 360°. In a particular embodiment, the field of view is up to 360° in a principal axis, which is often oriented to provide a 360° horizontal field of view. In this embodiment, a secondary axis may be defined, e.g., a vertical field of view. Such a vertical field of view may typically range from 0.1° to 180°, for example, from 1° to 170°. In accordance with the present invention, sections of the panoramic visual images may be selectively viewed. For example, while the panoramic visual images may comprise up to a 360° field of view, a smaller section may be selectively displayed, e.g., a field of from about 1° to about 60° may be selectively viewed.

As used herein, the term “directional sound” means the sound captured or reproduced to a listener as a function of a viewing direction selected from the panoramic visual images. In this manner, the directional sound is associated with the panoramic visual images. Preferably, the orientation of the directional sound corresponds with the viewing angle or selected section from the panoramic visual images. For example, multiple sound recording devices may be used to provide a virtual microphone during playback that can be pointed in the same direction as a virtual camera from selected panoramic data. In one embodiment, an estimate of distance between the camera and the sound source, e.g., either from processing the sound received at the camera or from an external source, can be used to co-locate sound with video at a point in space, rather than in just a particular direction.

An embodiment of the present invention provides an imaging system including a panoramic visual images display and a directional sound playback device. Examples of panoramic visual images displays include various types of computer monitors, televisions, video projection systems, head mounted displays, holograms and the like. The panoramic visual images display may comprise a single display device, or multiple display devices such as a row or array of devices. Examples of directional sound playback devices include one or more speakers driven by any suitable power source such as one or more amplifiers.

Fig. 1 is a schematic diagram illustrating a system 10 for producing panoramic images. A mirror 12 having an optical axis 14 gathers light 16 from all directions and redirects it to a camera 18. Visually, the immersive experience may be produced using a prerecorded or live sequence of images (possibly at TV frame rate but also at much slower frequencies) that are panoramic. A camera and panoramic mirror arrangement as shown in Fig. 1, or any other suitable panoramic imaging device, may be used to capture the panoramic visual images.

Fig. 2 is a raw image from a panospheric camera. Fig. 3 is the image from Fig. 2 displayed as a rectangular image using a projection onto a cylindrical surface. A viewer can select any part of this image to examine, e.g., as shown in the framed region. Each panoramic image may display a full 360° view from a point or a set of points. These images may be produced by any suitable type of panoramic camera, and may be viewed using any suitable projection (perspective, cylindrical, spherical, etc.) on a display device such as TV screens,

computer monitors, head mounted displays and the like. At any given time, only part of the image may be displayed to the user, based on commands given to the system.

In addition to panoramic imagery, sound may be recorded simultaneously. Sound may be captured on at least two channels and a temporal and spatial correspondence may be established between the panoramic images and the sound. Sound can be captured, for example, by any number of microphones each of which might be substantially uni-directional (only capture sound within a cone) or omni-directional (captures sound from all directions). Omni-directional microphones may be approximated by the use of several directional microphones placed in a ring. Sound may also be recorded separately, and an artificial coupling may be made between the panorama and the sound. It is also possible that either the panorama or the sound are synthetic, or both. That is, artificial panoramas created by computer models can be used in place of real panoramas and artificial sources of sound such as generated by a computer can be used, or a different sound that has been recorded separately may be associated with the video.

The sources of sound may be point sources (e.g., a singer on a stage) or a diffuse source (e.g., an applauding audience). The spatial correspondence between the panoramic images and sound can be achieved by localizing of the sources of the sound and embedding that information in the data stream that contains both the panoramic images and sound. The method for the localization of the sources of the sound may include measuring both the loudness and phase of the sound. From these measurements, an estimate of the location of the sources can be computed. If the panoramic images are generated using a rotating device, one rotating microphone can also be used to simulate two or more microphones.

In one embodiment of the invention as schematically illustrated in Fig. 4, a single panoramic camera is coupled with two or more microphones. The microphones may be located adjacent the camera, or may be located remotely from the camera. The audio recordings of the microphones may be synchronized with the video imagery. As the camera collects panoramic images, the sound is also captured by the microphones and the location of the source of the sound is computed. As the view is changed by the user manipulating a haptic

device such as joystick, the audio playback is altered such that the user is able to perceive the direction of the source of sound.

Fig. 5 schematically illustrates how the loudness of a sound from a particular sound source depends upon its location with respect to the current viewing direction. The sound at the upper right of Fig. 5 appears to come slightly from the left given the viewing direction shown. As the viewing direction changes, the strength of the sound can be made to vary as a function of angle between the viewing direction and the direction to the source of the sound. For the apparent angle  $\theta$ , one such function may be  $\cos(\theta/2)$ .

If the source of the sound remains constant, the user is able to sense the unvarying direction of the sound as the viewing direction is changed. Alternately, if the direction to the source of sound is changing, the user is able to sense the direction to the moving source as the selected view of the panoramic imagery is changed. One or more speakers can be used to playback the sound. If there is only one speaker, the loudness of the sound may be modulated according to the alignment of the sound source with respect to the current viewing angle. If multiple speakers are used for playback, the sound played back from the speakers may be modulated so as to provide the listener with the feedback.

Fig. 6 is the rectangular projected image of Fig. 3, illustrating an angle between the viewing direction (the center of the selected view) and the reference frame of the camera. Each row of the image spans  $180^\circ$ . If the viewing direction points to the location of the sound source, then the sound will be at its loudest. On the other hand, if the sound originates  $180^\circ$  from the viewing direction, the sound will be at its faintest.

If there is more than one speaker, the phase and loudness of the sound on those speakers may be modulated to emulate the position of the sound source with respect to the current viewing direction. Although there may be no depth information from the camera, the amount of zoom selected by the user could be interpreted as a depth cue to select the sound balance between the two microphones. It can also be used to alter the loudness of the sound so as to correspond with the experience of getting closer or farther away from the source of the sound. Thus, sound may be recreated as coming from a direction without knowing its exact position.

If the directional microphones are not at the optical center of the panoramic device, using the angular difference between the viewing direction and the microphone direction may not be sufficient. For example, three omni-directional microphones may be placed in an environment as illustrated in Fig. 7. A listening distance may be recreated by modulating the sound from the two microphones as a function of the angular difference between the viewing direction and the axis of the microphone baseline. In Fig. 7,  $b_{12}$  is the baseline distance between microphones 1 and 2, while  $b_{23}$  is the distance between microphones 2 and 3. The angle  $\theta$  is the angle between the baseline and the viewing direction.

Only those microphones that fall in the field of view may be used to recreate the sound. In the embodiment shown in Fig. 7, microphones 1 and 2 are used while microphone 3 is not used. If the field of view were to rotate clockwise, microphone 1 would not be used but microphones 2 and 3 would be used. The sound is composed based on combinations of the sound recorded at two microphones. This may be done for every pair of microphones in the viewing area. The strengths of the sounds from the microphones may be combined as follows. The relative strength of the signal of microphone  $i$  is given by:

$$\frac{b_{ij} - d_i}{b_{ij}}$$

where  $b_{ij}$  is the baseline between microphones  $i$  and  $j$  and  $d_i$  is the distance between microphone  $i$  and the intersection of the axis of the viewing direction and  $b_{ij}$ . The effect of direction (the offset  $\theta$ ) may be computed as illustrated in Fig. 5.

In another embodiment of the invention shown in Fig. 8, two simultaneously recorded sequences of panoramic imagery may be played back such that the user is able to perceive depth in the selected view in addition to the sound as in the previous embodiment. The depth may be either directly inferred by the user viewing the multiple image streams, or may be based on a 3D model that is extracted by a computer process that finds correspondence between features in the multiple views or tracks the image features in one or more image sequence to create the three dimensional model.

In a preferred embodiment of the invention, one panoramic camera may be used in conjunction with multiple microphones. The microphones have directionality, i.e., they are sensitive to sounds coming from the direction that they are pointed, with sensitivity falling off

in other directions. The microphones may have overlapping fields of sensitivity. Any sound in the environment may be detected by at least two microphones. Sounds from the environment are recorded simultaneously with the video and can be correlated to the video for direct transmission or playback. The camera may have a natural frame of reference and sounds may be located either by position or direction (or both) with respect to the frame of reference. When the panoramic image is unwarped, the direction that the viewer chooses defines the offset from the camera reference frame. This offset may change dynamically with the selected view. The signal recorded from each microphone may be played back in a modified manner based on the offset from the camera reference frame. Sound from each microphone may be composed depending on the number of playback devices. If only one speaker is available, then the sounds recorded from all microphones may be simply added up.

If the offset between the direction of the microphone  $i$  and the camera reference is denoted by  $\theta_i$ , the strength of the signal associated with that microphone,  $M_i$ , is  $\cos(\theta_i) + \varepsilon$ , where  $\varepsilon$  is a minimal level of sound playback. The composite sound is created by:

$$\sum (\cos(\theta_i) + \varepsilon) \cdot M_i$$

If the playback device consists of multiple speakers, the sound may be distributed to each speaker such that each speaker only plays the sounds corresponding to microphones pointed in a certain sector. For example, if four speakers are used, each speaker may only play sounds attributed to microphones in a 90 degree sector.

The present invention may be used for various telepresence applications that involve capturing of an event. Some examples of the possible applications include entertainment, surveillance and tourism.

Whereas particular embodiments of this invention have been described above for purposes of illustration, it will be evident to those skilled in the art that numerous variations of the details of the present invention may be made without departing from the invention as defined in the appended claims.